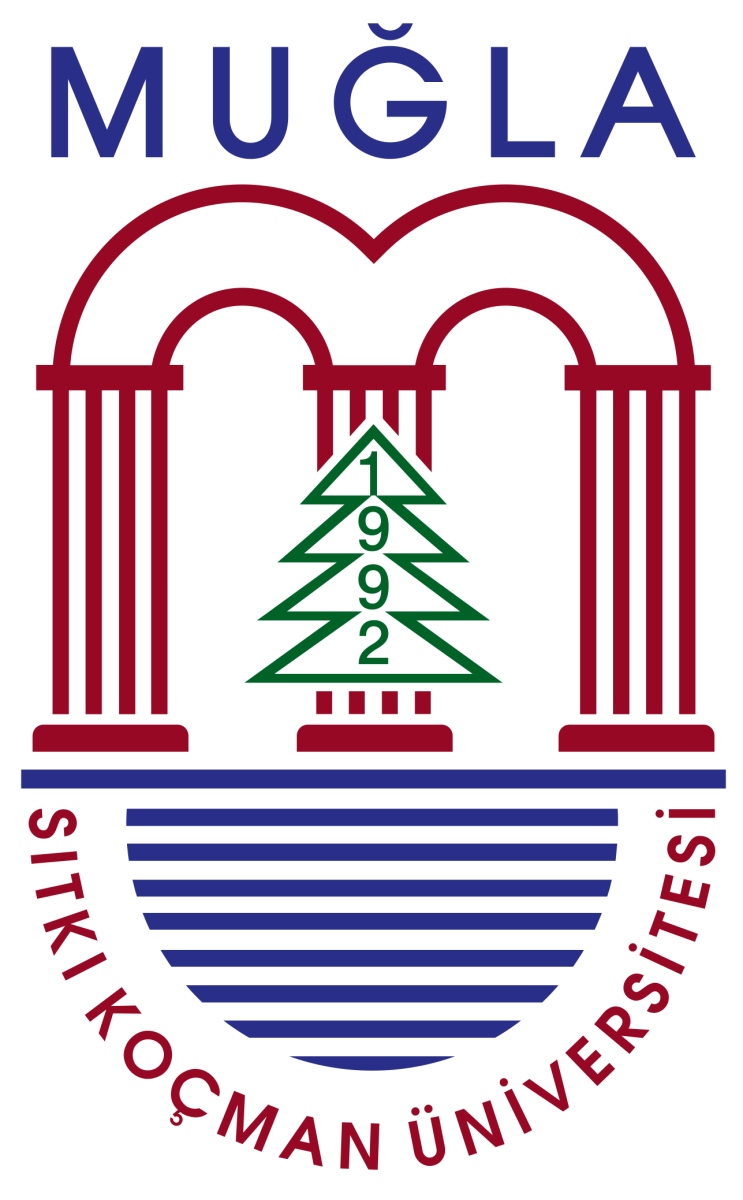
**Clinical Support Mobile Application Development for Lung Cancer Diagnosis**



**Computer Engineering - Senior Project Report**

**Advisor :** Tuğba SÜZEK, PhD

**Group Members :** Emre Taşkın

**Date :** 25/01/22

***Clinical Support Mobile Application Development for Lung Cancer Diagnosis***

Emre Taşkın

Computer Engineer

Muğla Sıtkı Koçman University

25/01/22

**Summary**

Lung cancer (both small cell and non-small cell) is the second most common cancer in both men and women.[1] Lung cancer is divided into two main classes as Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC). NSCLC type lung cancer constitutes approximately 80-85% of lung cancers.[2] The types that will be included in the data set that we will use during our research are Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC), which are two of the three subtypes of Non-Small Cell Lung Cancer (NSCLC), as well as Small Cell Lung Cancer (SCLC). ) type.

The interpretation of pathology specialists about which subtype of lung cancer is found in the patient as a result of the test is important in treatment planning. In some cases, pathologists may have difficulty in making a definitive diagnosis about the test result and may prefer to reach a common decision with other teammates. In such cases, thanks to the clinical support mobile application we will develop, the tissue images transferred to the digital environment will be analyzed by machine learning methods and the pathology specialist will be assisted in making a diagnosis. The most unique aspect of our proposal is the absence of a mobile application that is currently trained in the clinic with histopathological images of lung cancer that is available to the public and where the clinician can get decision support using their own mobile phone.

**Table of Contents**

[1. Introduction](#_gjdgxs) 3

[2. Methods](#_30j0zll) 3

1. Creation and training of machine learning model3
2. Mobile Application Development 34

3[. Conclusion](#_3rdcrjn) 5

4[. References](https://docs.google.com/document/d/1SUeU3gjQgP9Wf1nlIwy1bismFiAFqziKon6DlzdaCA8/edit#heading=h.1ksv4uv) 5

# **Introduction**

Clinical management of patients with LUAD (adenocarcinoma), LUSC (squamous cell), or SCLC (small cell) lung cancer is a multi-clinician-led process. The diagnosis process begins with the referral of the patient to the Radiology Specialist from the Chest Diseases specialist, whom the patient reaches with a complaint, and then the referral to the General Surgery specialist for surgical intervention. The treatment process begins when the General Surgery specialist directs the patient to the Medical Oncology specialist, together with the diagnosis made by the Pathology specialist. The diagnosis of the pathologist is a critical step in the treatment, as the lung cancer subtype (LUAD, LUSC or SCLC) to be treated is determined according to the pathology result in lung cancer. In order to minimize human errors in this multi-clinician process, clinical support software that pathologists can get support with easy access at the points where they are hesitant in diagnosis is critical for the correct planning of the remaining treatment process of the cancer patient.

# **Methods**

When our project is considered from a macro perspective, it consists of two main stages:

## 

## **Creation and training of machine learning model**

The Cancer Genome Atlas network (TCGA)[3], which is open for use and includes cancer data, includes data from approximately 11000 patients). Lung cancer image data belonging to 3 subtypes in TCGA (LUAD, LUSC and SCLC) will be downloaded using the TCGABiolinks[4] package in the R programming language and will be cleaned with the help of the HistoQC[5] package in order to obtain an efficient training model. We will divide our cleaned dataset into two sets, training and testing, without disturbing the response variable balance. The machine learning model will be trained with the help of the images in the training set with the help of the CNN algorithm using the appropriate parameters.

1. **Mobile Application Development**

After the machine learning model is created, the model learned using Chaquopy[6] will be combined with a mobile application to be developed for Android devices in order to serve its purpose. A simple and easy-to-understand interface will be designed so that it is easy for clinicians to use in daily practice.set with the help of the CNN algorithm using the appropriate parameters.

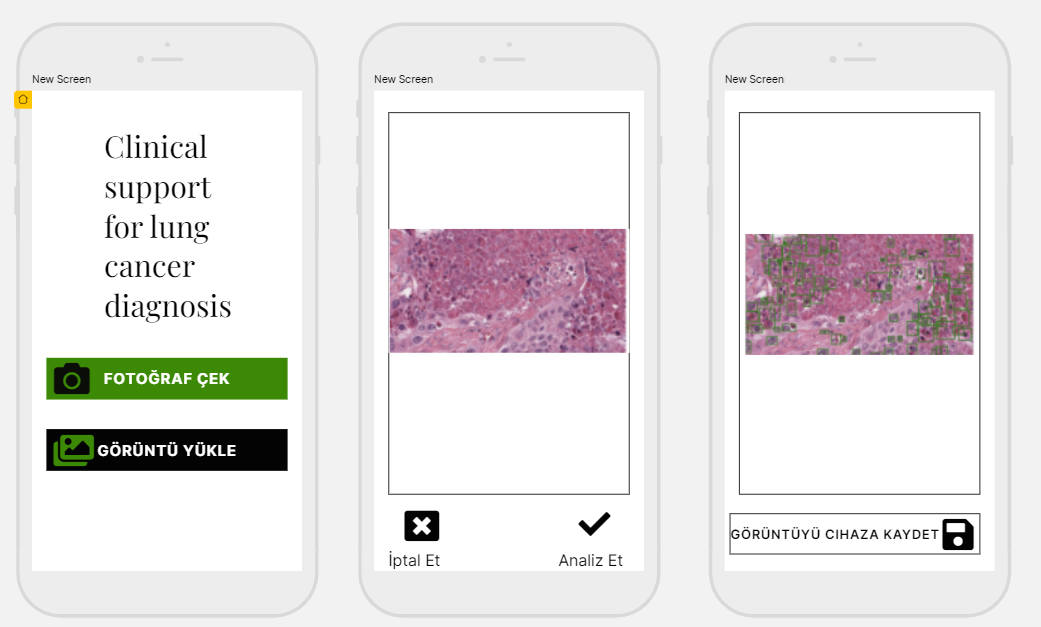
Our mobile application, Pathology specialist clinicians in Muğla Sıtkı Koçman Faculty of Medicine will be contacted and feedback will be received in order to make the user interface user-friendly. If our application is found successful by the clinician, our end user, an ethics committee application will be made to Muğla Sıtkı Koçman University human research ethics committee for testing the system with 20 real histopathology data. With the help of the feedback received after the testing phase, the software will be finalized and made ready for use.

# **Results**

In first semester, I have completed AI for medical diagnosis course, I learned to use TCGABiolinks library in R language and I planned the interface design of the mobile application.

First of all, I installed all the libraries required to use TCGABiolinks on my computer. Then I ran the required code to download the data. At this point I ran into memory and internet speed issues. The file size of the image data I used was too big for my computer. Downloading images was also taking too long. I was able to download a very small portion of my total data size due to memory problem. I downloaded HistoQC and searched for usage examples but couldn't find it, I'm still researching how to use it to get rid of dirty data.

Secondly, I designed prototype for the mobile application using an online platform.



***Figure -1:*** Mobile Application Design

In second semester, I will complete remaining parts. I will try to reach high scores by training my artificial intelligence model with the Convolutional Neural Networks algorithm.

Then I will develop a mobile application for android devices using Android Studio and combine my model with the application using Chaquopy. As the final stage, it is planned to publish the mobile application on Google Play and prepare an oral statement after real-time testing on test images downloaded from TCGA.

# **Conclusion**

Thanks to the mobile application we will develop to support pathology specialists, pathologists will be able to both be sure of the accuracy of the results they interpret and detect an overlooked phenomenon. Pathologists will introduce the images they have obtained to the application through the phone camera or uploading the image on the device, and the machine learning algorithm that will run in the background will analyze the uploaded image.

# **References**

[1]Cancer.org. 2020. <https://www.cancer.org/content/dam/CRC/PDF/Public/8703.00.pdf>

[2]Cancer.org. 2019. What Is Lung Cancer? | Types Of Lung Cancer. <https://www.cancer.org/cancer/lung-cancer/about/what-is.html>

[3]The Cancer Genome Atlas Program

<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>

[4] TCGAbiolinks: An R/Bioconductor package for integrative analysis with GDC data

<https://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html>

[5] "HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides", Janowczyk A., Zuo R.,

Gilmore H., Feldman M., Madabhushi A., JCO Clinical Cancer Informatics, 2019

[6] Chaquopy Python SDK for Android <https://chaquo.com/chaquopy/>